Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 93 (2016) 503 – 512

Procedia Computer Science

6th International Conference On Advances In Computing & Communications, ICACC 2016, 6-8 September 2016, Cochin, India

Robust Face Recognition System in Video using Hybrid Scale Invariant Feature Transform

Mohanraj. V[a*], Vimalkumar. M[b], Mithila. M[c], Vaidehi. V[d]

[b,c]*Department of Information Technology, Madras Institute of Technology, Anna University, Chennai-44.*
[a,d]*Department of Electronics Enigneering, Madras Institute of Technology, Anna University, Chennai-44.*

Abstract

Face recognition plays a significant role in the research field of biometric and computer vision. The important goal of an efficient Face Recognition (FR) system is to have negligible misclassification rate. In video-based face recognition system, the illumination and pose variation problems are predominant. Most of the efficient FR systems are developed for controlled or indoor environment, hence they fail to give accurate recognition in outdoor environment of different illumination variation. Other challenges include occlusion and facial expression. The illumination problem is handled by Histogram Equalization in existing methods. The original Scale Invariant Feature Transform (SIFT) also works well only for pose variation and fails to produce satisfactory results under varying illumination. Hence Hybrid Scale Invariant Feature Transform (HSIFT) with Weighting Factor in feature matching is proposed in this paper which uses a fixed facial landmark localization technique and orientation assignment of SIFT to extract illumination and pose invariant features. The extracted features are then matched using Fast Library for Approximation of Nearest Neighbor (FLANN). The proposed method has been implemented in OpenCV to give a recognition rate of 98% and 95.5% in YouTube celebrity and Extended Yale B dataset respectively.

Peer-review under responsibility of the Organizing Committee of ICACC 2016

Keywords: Face recognition; SIFT; Ensemble of regression trees; Weighting factor;

1. Introduction

Face recognition is one of the fastest growing research areas, owing to its significance in security surveillance, building/store access control and several other applications. It is an application used to detect, identify

* Corresponding author. Tel.: +91-9944257261.
E-mail address: mohanraj4072@gmail.com

Peer-review under responsibility of the Organizing Committee of ICACC 2016
doi:10.1016/j.procs.2016.07.240

and recognize human faces from video or images. Face is the most distinctive region of human and is very hard to forge, hence it plays a vital role in security applications. Some of the challenges in face recognition are illumination, pose, occlusion, ageing, expressions and low-resolution. There are several works that deal with each of this challenge, in the proposed work the focus is illumination and pose. Most of the applications of face recognition are in uncontrolled environments and in such cases illumination and pose are the most challenging factors to overcome. Face detection is the first step in recognition as only after the face region is detected, features can be extracted and recognized. Face detection techniques based on Haar and HoG[1, 2] have several advantages and disadvantages. HoG is used in the proposed work as it outperforms the other methods in terms of detection accuracy. It has high detection and low false positive rates which makes it robust. In order to overcome illumination variations and shadowing it performs contrast normalizations.

There are two major approaches for feature extraction, typically holistic feature and local feature. In holistic feature based approach like Eigen face the features are extracted from the face as a whole which may sometimes be affected by occlusion and expression changes. Whereas in local features based approach, these issues are overcome as only patches of the image are considered, also they are scale and rotation invariant. There are several local feature extractors like Gabor, LBP, SIFT, SURF etc that are found to be effective, yet fail in certain conditions. For instance, SIFT is scale and rotation invariant as it is based on local features, but it suffers from varying illumination conditions. Similarly each feature extraction method fails in handling more than one challenge; hence there is a need for a hybrid feature extractor. In the proposed work, SIFT is adopted and modified by including fixed landmark points and light adaptation filter based on retina modelling to overcome illumination changes, giving better recognition rates when tested in real-time video datasets as well as standard datasets such as YouTube celebrity. The paper is organized as follows; section 2 consists of related works, section 3 explains video based Face Recognition system, section 4 shows the Implementation and Results and section 5 gives the conclusion and future work.

2. Related Works

CemilTosik et al.[1] proposed an Illumination Invariant Face Recognition system. In their system, Histogram equalization, Discrete Cosine Transforms (DCT) and steerable Gaussian filters are applied to face images as a pre-processing technique. It is found that Histogram equalization with Steerable Gaussian filters gives the best performance under varying illuminations. Ngoc-Son Vu et al.[2] proposed an Illumination-robust face recognition. Illumination normalization in their work is based on retina modelling, which combines two adaptive nonlinear functions and a Difference of Gaussians filter. It achieves very high recognition rates even for the most challenging illumination conditions. Paul Viola and Michael Jones[3] in their work have described a machine learning algorithm for object detection called “Rapid Object Detection using a Boosted Cascade of Simple Features” which is capable of rapid processing and high detection rates. This detector used in real-time applications, runs at 15 frames per second without getting affected to skin color detection. However a few misclassifications are found in certain lighting conditions. Navneet Dalal and Bill Triggs[4] proposed Histograms of Oriented Gradients for Human Detection. Their work is based on the idea that the objects appearance and shape can be characterized by the local intensity gradients and edge directions. Using this type of locally normalized Histogram of Oriented Gradient features gives very good results and is unlike the well-known Haar object detection algorithm which has many false positive rates.

Shaoqing Ren et al.[5] proposed a method of Face Alignment via Regressing Local Binary Features in which two methods are introduced such as local binary features and a method for learning those features. It is observed that their method is computationally less complex and achieves over 3,000 fps for localizing the landmarks. Vahid Kazemi et a1.[6] proposed an one millisecond face alignment with Ensemble of Regression Trees that localize facial landmark positions directly from pixel intensities, achieving good performance and quality predictions in real time. David G. Lowe[7] has proposed Distinctive Image Features from Scale-Invariant Key points that are invariant to rotation, scale and change in viewpoint as it is based on the local features rather than the holistic image. These features can generally be used for object recognition and in the proposed work it is used for face recognition. Renliang Weng et al.[8] proposed a Robust Point Set Matching for Partial Face Recognition. Generally holistic facial

features are used in face recognition. However in an unconstrained environment obtaining such features become difficult as the face may be occluded by other objects. This problem is overcome in their work by a new partial face recognition technique to recognize people from their partial faces. Given a pair of gallery image and probe face patch, first they detect key points and extract their local textural features. Then, a Robust Point Set Matching (RPSM) method given by Renliang[9] is used to discriminatively match these two extracted local feature sets, where both the textural and geometrical information of local features are explicitly used for matching simultaneously. Though there are several face recognition techniques available in the literature, the face recognition in video stream demands a robust and efficient scheme under different illumination, pose and expression variation.

3. Proposed HSIFT scheme for Video Based Face Recognition system

Video obtained from surveillance camera in uncontrolled environment produce images of varying illumination and poses. The proposed HSIFT scheme based face recognition involves Retina Modelling[2], HOG descriptor based Face detection[4], Landmark localization upon detected faces using Ensemble of Regressors[6], Feature extraction from the localized facial key points[8] and feature matching for final face recognition process using Fast Library Approximate Nearest Neighbour (FLANN).

A Face Recognition model is developed by modelling the every process involved in the Face Recognition system as presented in the following equations. The input video stream is pre-processed through two stages of Non-Linear adaptive Filter as given in Equations 1 and 2.

$$I_{la1} = (\max(I) + F_1(p))\frac{I(p)}{I(p)+F_1(p)} \quad (1)$$

$$I_{la2} = (\max(I_{la1}) + F_2(p))\frac{I_{la1}}{I_{la1}+F_1(p)} \quad (2)$$

where,

$$F_1(p) = I(p) * G_1 + \frac{\bar{I}}{2}$$

$$F_2(p) = I_{la1}(p) * G_2 + \frac{\bar{I}_{la1}}{2}$$

$$G_1(x, y) = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\sigma_1^2}}$$

$$G_2(x, y) = \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\sigma_2^2}}$$

The pre-processed video is then applied through the HOG descriptor face detection chain to detect faces from the video stream. The HOG face detector is given in Equation 3. The set of faces detected is collectively represented $\hat{P}$ as given in Equation 4.

$$\hat{P} = I_{la2} * S \quad (3)$$

$$\hat{P} = (L_1, L_2, \ldots\ldots L_n) \quad (4)$$

where,

$L_1, L_2, \ldots\ldots L_n$ – Number of detected faces

$$S = \tan^{-1}\left(\frac{L_x}{L_y}\right)$$

The landmarks from the detected faces are localized based on a defined shape template presented in Equation 5. The initial shape template is applied upon the detected face and approximated to identify the current shape estimate from every detected face accurately as shown in Equation 6.

$$S = (X_1^T, X_2^T, X_3^T, \ldots\ldots\ldots X_p^T) \quad (5)$$

$$S^{(t+1)} = S^{(t)} + r_t(L_n, S^{(t)}) \quad (6)$$

where,

$X_1^T, X_2^T, X_3^T, \ldots\ldots\ldots X_p^T$ – denotes the coordinates of all the p facial landmarks in I

$S^{(t)}$ – Current shape estimate

From the localized facial keypoints SIFT feature descriptors are extracted by computing the Magnitude M(x, y) and orientation O(x. y) by using the Equations 7 and 8.

$$M(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))2 + (L(x,y+1) - L(x,y-1))2} \quad (7)$$

$$O(x,y) = \tan - 1((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \quad (8)$$

The final facial feature vector is the combined output of the independently computed magnitude and orientation values as shown in Equation 9.

$$T = M(x, y) * O(x, y) \quad (9)$$

where,

T – Feature Vector.

The proposed HSIFT scheme tries to overcome the problems caused due to illumination and pose variation in video using Hybrid Scale Invariant Feature Transform based Face Recognition system. SIFT feature extractor is modified to overcome this illumination and pose variation problem using a fixed landmark localizer called Ensemble of regression Trees and retina modeling to produce an illumination insensitive representation of an input face image. The features thus obtained are matched using Euclidean distance. Weighting factor is applied to the prominent features to improvise the FR results. Figure 1 presents the architecture of the proposed Face Recognition system. The algorithm Hybrid SIFT with weighting factor shows each steps involved in the proposed Face Recognition system.

3.1. Preprocessing

Retina modeling[2] imitates the process of human retina. Human retina has the natural ability that enables human to see objects in different illumination conditions. Retina modelling combines two adaptive nonlinear functions and a Difference of Gaussians (DoG) filter, which is similar to the performance of two layers of the retina: the photoreceptors and the outer plexiform layer. In this work, only the two non-linear functions are used, DoG is neglected.

3.2. Face Detection

HoG[3] feature descriptors are also used for object detection. HoG works based on the concept that the local object appearance and shape can be determined by its edge directions or the distribution of local intensity gradients. The image is divided into regions called cells and a local 1-D histogram of gradient directions or edge orientations is built for each cell. The cells are normalized by accumulating a measure of local histogram “energy” into bigger regions called blocks. This normalized descriptor blocks are referred as Histogram of Oriented Gradient (HOG) descriptors. A human detection chain is then generated by overlaying the detection window with a grid of HOG descriptors and with the combined feature vector in a SVM based window classifier.

3.3. Feature Extraction

In this paper, a fixed landmark localizer is used to localize the prominent landmarks on the face. Illumination and pose invariant feature vectors are then extracted from these key points by computing orientation and magnitude based on the SIFT feature extraction algorithm.

3.3.1. Construction of Feature Vector

Landmark localization has always been inhibited by occlusion and pose variation. Facial landmark localization of eyes, nose and mouth are vital for tasks like face recognition, face tracking, face animation and 3D face modeling. But identifying these landmarks are challenging due to large variations on facial appearance, illumination, and partial occlusions. Facial landmarks in this paper are localized by an “Ensemble of Regression Trees” technique given by Vahid Kazemi et al[6].

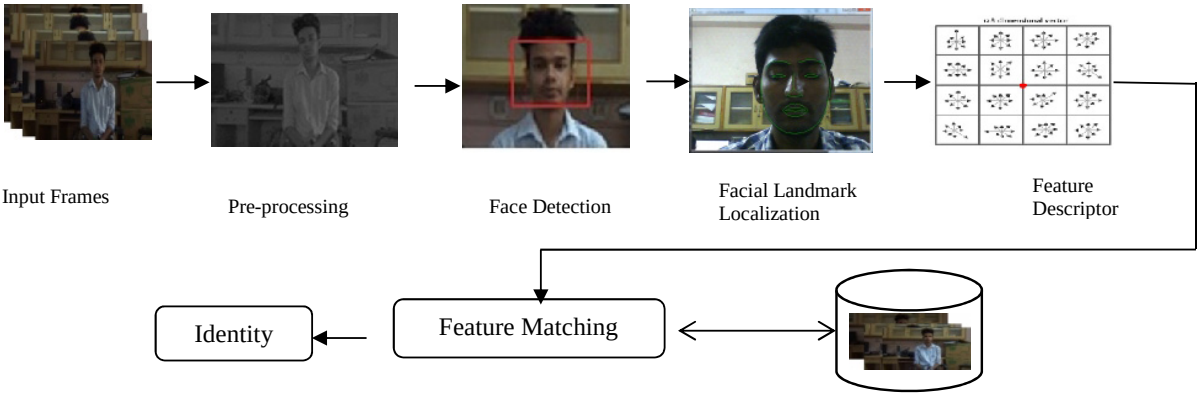


Fig. 1 Architecture of proposed Face Recognition system.

Table 1. Algorithm for Hybrid SIFT with weighting factor

INPUT : Video frames (f)
OUTPUT: Recognized result
Begin
1. Apply retina filter R for each frame: $R(f) = (Iin(max) + F1(p))\, Iin(p) / Iin(p) + F1(p)$.
2. Detect Faces using Histogram of Orientated Gradients (HoG).
3. Extract Feature
for f=1…..N:
Set of keypoints are detected using the ensemble of regression trees
$rt(I, S\hat{}(t)) = fk - 1(I, S\hat{}(t)) + \nu\, gk(I, S\hat{}(t))$
for each keypoint (x, y) :
construct the 16x16 neighborhood, and divide into 4x4 sub region
$M(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))2 + (L(x,y+1) - L(x,y-1))2}$
$O(x,y) = \tan - 1((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$
The O(x, y) is put into a 8 bin histogram and a 128 feature vector (4*4*8)is constructed.
4. Apply Weighting factor (w_i) to all the landmarks with more weightage to eyes compared to nose and mouth.
5. Match Gallery and probe image features using the fast approximate nearest neighbors.
6. for f_j = 1……..K
if (f_j> th)
Result: Known Face
else
Result: Unknown Face
Endif
End

Iin= input image
G1(*x*, *y*) =(1/2 $_1$2)$e^{-x2+y2/2}$ $_1^2$
Sˆ(t)=vector of keypoints
th=minimum threshold for matching
L(x,y) – Intensity of the pixel.
F1(*p*) = Iin(*p*) ∗ G1 + Iin/2
N: no of detected faces
K=size of db
x,y – co-ordinate points.

3.3.2. Construction of Feature Vector

SIFT feature extraction algorithm given by David G Lowe[2] is scale and rotation invariant as it is based on local features. Apart from this, these features are robust to variations in viewpoint as well as addition of noise and hence this algorithm can be used for recognition purposes. SIFT has four key steps that are based on two major ideas, the first one involves identifying the keypoints and the second is about constructing the feature vector for the determined keypoints. Since the keypoint localization in SIFT is affected by different lighting conditions and noise, in this paper only the orientation assignment step of SIFT is used and the keypoint localization step is modified by using an Ensemble of Regression Trees as explained above to localize the key points.

Every keypoint is assigned an orientation based on the local image properties that make it invariant to image rotation. For each keypoint of an image L(x, y), the magnitude m(x, y) and orientation O(x, y) is computed using the formula (10) and (11).

Gradient Magnitude formula:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))2 + (L(x,y+1) - L(x,y-1))2} \quad (10)$$

Orientation formula:

$$O(x,y) = \tan - 1((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \quad (11)$$

After orientation assignment, for the local image region, a descriptor is computed around the keypoint. A 16x16 window which is broken into sixteen 4x4 sub-windows is constructed around the keypoint. For each pixel in this 4x4 sub-window an 8 bin histogram is constructed. For a single 4x4 region the 16 random orientations are put into a fixed 8 bin histogram. Similarly this is done for all the other 4x4 regions and a 4x4x8(128) feature vector that uniquely identifies the keypoint is obtained.

3.4. Descriptor Matching

Recognition is achieved by matching the query and trained features using some nearest neighbour algorithm. The database consists of features of the trained images in a sequential manner representing a person for a particular range. For example the first 20 features in the database represent person 1, the next 20 represent the next person and so on. The query image's features obtained are compared with the database image features using Euclidean distance formula.

Euclidean Distance:

$$d((x1,y1)(x2,y2)) = \sqrt{(x1-x2)2 + (y1-y2)} \quad (12)$$

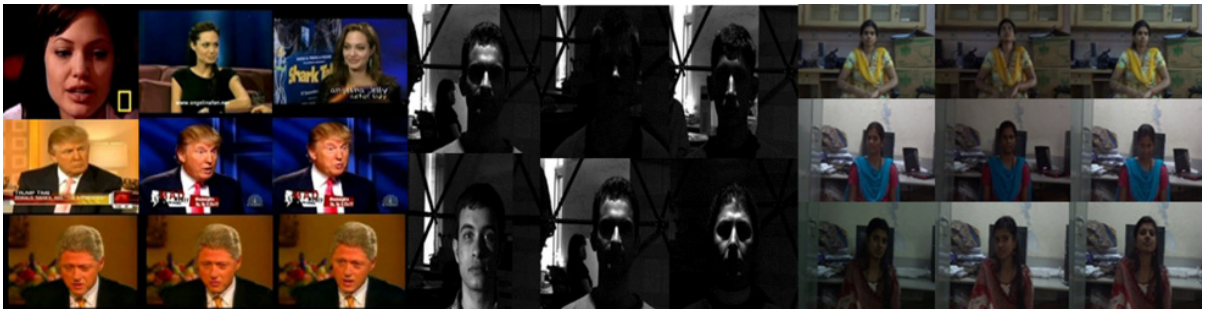
26 feature vectors obtained from the landmark points of the query image is matched across the database features, the range having the highest similar features based on the distance matching algorithm is considered as the result. In some cases misclassifications may occur, therefore in order to improvise the performance of the overall system weighting factor is applied in this work. Through experimentation, it is found that the nose, nose tip are a more prominent feature than the eyes and eyebrows. Hence a weight of 4,3,2,1 is given to the nose, nose tip, eyes and eyebrows respectively, which means that if a nose feature is matched to the index pointing to it is given a plus 4 weightage whereas for an eyebrow it has the same value. In this manner the number of features matched is counted and the range within which they fall is calculated. The range having the maximum number of matches is declared as the recognised person. This way a number of misclassifications can be reduced and the recognition rate can be improvised.

4. Implementation and Results

The implementation is done in a 64-bit windows operating system with 8-GB RAM and INTEL i5 processor chip. The visual studio 12 IDE and OpenCV 2.4.9 is used to compile and run the project. C++ programming language is used for implementing the proposed system.

4.1. Youtube Celebrity Dataset

The youtube celebrity database[19] is a standard video database that consists of 3,452 videos of 1,595 people. Each person has videos in about three to four different environments, thereby giving different pose and illumination variations. For this paper a subset of videos from the YouTube celebrity database is taken. 30 people are trained and 20 frames for each person are registered by utilizing two videos for each of them. Some of the frames obtained from the video used for training are shown in Figure 2a. The database size is 600. For testing, the remaining videos are used to correctly identify the person.



(a) (b) (c)

Fig. 2. (a). Sample frames from YouTube Celebrity dataset (b). Sample images from Extended Yale B dataset (c). Sample frames from MIT_INDIA dataset

4.2. Extended YaleB Dataset

The Extended Yale B database consists of 28 human subjects under 9 poses and 64 illumination conditions leading to a total of 16128 images. The data format of the images in Extended Yale B is the same as in Yale B database. From the Extended Yale B database, 21 people are trained and for each of them five images under varying illuminations are used. Some of the images used for training are shown in Figure 2b.

4.3. MIT_INDIA Dataset

The MIT INDIA database is a real-time video database consists of 20 human subjects under 15 different poses and 2 illumination conditions. For training, 15 images of same subject under two illumination conditions are taken. Some of the images used for training are shown in Figure 2c.

4.4. Performance measure on Youtube celebrity Dataset

For YouTube celebrity database two different scenarios are considered. In the first scenario, the subjects are trained and tested with same background but different poses. In the second scenario, the subjects are tested with different background, illumination conditions and poses. From Table 2 it can be seen that the accuracy of YouTube celebrity database in same scenario for range summation, weighting factor and both range summation and weighting factor is 98%, 98% and 97.5% respectively. Similarly, the misclassification rate is 1.5%, 2.0% and 0.5% respectively. Though the accuracy of range summation and weighting factor is higher than both combined, the misclassification rate is reduced when both range summation and weighting factor are combined.

Table 2.Face Recognition (FR) results on YouTube celebrity database in same scenario

Proposed Method	Range summation	Weighting factor	Range summation and weighting factor
Accuracy	98.0%	98.0%	97.5%
Misclassification Rate	1.5%	2.0%	0.5%

From Figure 3 it can be seen that the accuracy of YouTube celebrity database in different scenario for range summation, weighting factor and both range summation and weighting factor combined is 73.75%, 80.62% and 78.7% respectively. Similarly, the misclassification rate is 26.2%, 17.5% and 9.37% respectively. The accuracy and misclassification rate of both range summation and weighting factor combined is comparatively good than the other two methods.

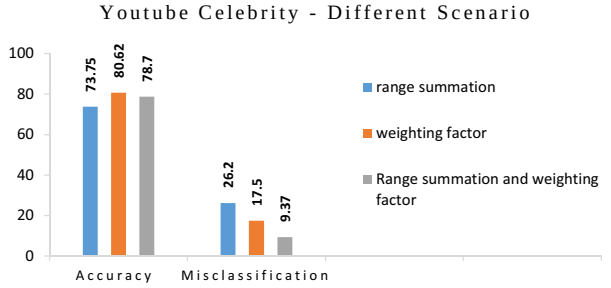


Fig. 3. FR results on YouTube celebrity database in different scenario

4.5. Performance measure on Extended YaleB Dataset

For testing, 42 images of the subjects with different illumination condition and pose variations are taken. An accuracy of 95.23% is achieved as shown in Table 3.

Table 3.Face recognition results on Extended Yale B database.

Database	Extended Yale B database
Number of Subjects	21
Database Size	105
Number of images for testing	42
Accuracy	95.23%

4.6. Performance measure in MIT INDIA Dataset

For MIT_INDIA database two different illumination conditions and 15 pose variations are considered. 20 subjects are tested in real-time and Retina modelling is used for pre-processing the frames. An accuracy of 84.86% is achieved as shown in Table 4.

Table 4. Face recognition results on MIT_INDIA database.

Database	MIT_INDIA database
Number of Subjects	20
Database Size	300
Accuracy	84.86%

4.7. Comparison on Video based Face Recognition results on YouTube Celebrity database

The proposed HSIFT method gives better results when compared to other video based face recognition methods on YouTube Celebrity database as shown in Table 5.

Table 5. Comparison on Video based Face Recognition results on YouTube Celebrity database.

Method	Accuracy
MSM[11]	61.1
MMD[12]	62.9
MDA[13]	65.3
CHISD[14]	66.3
SANP[15]	68.4
COV+PLS[16]	70.1
MA[17]	74.6
MSSRC[18]	80.8
IP/EPD[19] (Same environment)	81.9
IP/EPD[19] (Cross environment)	78.6
Proposed HSIFT method (Same environment)	98.0
Proposed HSIFT method (Cross environment)	80.6

5. Conclusion and Future Work

This paper proposed Hybrid SIFT scheme to recognize face in video under different illumination and pose variation. The proposed system uses light adaptation filter of Retina Modelling for pre-processing and Histogram of Oriented Gradients for face detection, as it has lesser false positive rates. In order to overcome the illumination problem, the original SIFT is modified in this paper with fixed landmark localizer to mark the key points from which the orientation assignment of SIFT is carried out and an illumination invariant feature vector is obtained. The features are matched using Fast Approximate Nearest Neighbor search library (FLANN) and weightage factor is applied to improvise the accuracy of the system. The system has been implemented and validated with real time and benchmark databases such as YouTube celebrity and Extended Yale B databases. It is found from experimental results that the proposed method gives accurate result and takes lesser time for face recognition. In future, the proposed method can be validated in a larger database and the recognition time can be reduced using clustering algorithms. The accuracy of the FR system can be improved by deep learning techniques such as Convolution Neural Network (CNN).

Acknowledgements

This research project is supported by DAE-BRNS, Department of Atomic Energy, Government of India. The authors would like to extend their sincere thanks to DAE-BRNS for their support.

References

1. Cemil Tosik, Alaa Eleyan, Mohammad Shukri Salman. *A Illumination Invariant Face Recognition System.* 978-1-4673-5563-6/13/$31.00 ©2013 IEEE
2. Ngoc-Son Vu, Alice Caplier. *An Illumination-robust face recognition using retina modelling*. 978-1-4244-5654-3/09/$26.00 ©2009 IEEE

3. Paul Viola, Micheal Jones, *Rapid object detection using a boosted cascade of simple features.* Conference on Computer Vision and Pattern Recognition, 2001.
4. Navneet Dalal and Bill Triggs. *Histograms of Oriented Gradients for Human Detection.* CVPR05.
5. Shaoqing Ren, Xudong Cao, Yichen Wei Jian Sun. *Face Alignment at 3000 FPS via Regressing Local Binary Features.* Computer Vision Foundation, 2014
6. Vahid Kazemi, Josephine Sullivan, *One Millisecond Face Alignment with an Ensemble of Regression Trees.* Computer Vision Foundation, 2014
7. Dong Li, Huiling Zhou and Kin-Man Lam. *High-Resolution Face Verification Using Pore-Scale Facial Features.* IEEE Transactions on image processing, VOL. 24, NO. 8, August 2015.
8. David G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints.* International Journal of Computer Vision, 2004.
9. Renliang Weng, Jiwen Lu and and Yap-Peng Tan. *Robust Point Set Matching for Partial Face Recognition.* IEEE Transactions on Image Processing 2015.
10. Yamaguchi, K. Fukui, and K. Maeda. *Face recognition using temporal image sequence*. In IEEE International Conference on Automatic Face and Gesture Recognition, pages 318–323, April 1998.
11. R. Wang, S. Shan, X. Chen, and G. Wen. *Manifold-manifold distance with application to face recognition based on image set*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, June 2008.
12. R. Wang and X. Chen. *Manifold discriminant analysis*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 429–436, June 2009.
13. H. Cevikalp and B. Triggs. *Face recognition based on image sets*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2567–2573, June 2010.
14. Y. Hu, A. S. Mian, and R. Owens. *Sparse approximated nearest points for image set classification*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 121–128, June 2011.
15. R. Wang, H. Guo, L. Davis, and Q. Dai. *Covariance discriminative learning: a natural and efficient approach to image set classification*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2496–2503, June 2012.
16. Z. Cui, S. Shan, H. Zhang, S. Lao, and X. Chen. *Image sets alignment for video-based face recognition*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2626–2633, June 2012.
17. E. G. Ortiz, A. Wright, and M. Shah. *Face recognition in movie trailers via mean sequence sparse representation-based classification*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 3531–3538, June 2013.
18. Ming Du and Rama Chellappa. *Video-Based Face Recognition Using the Intra-Personal/Extra-Personal Difference Dictionary*. Proceedings of the British Machine Vision Conference. BMVA Press, September 2014.
19. M.Y. Kim, S. Kumar, V. Pavlovic, and H.A. Rowley. *Face tracking and recognition with visual constraints in real-world videos*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, June 2008.